Detection of haemagglutinin D222 polymorphisms in influenza A(H1N1)pdm09-infected patients by ultra-deep pyrosequencing

M. Selleri[1], A. Piralla[2], G. Rozera[1], E. Giombini[1], B. Bartolini[1], I. Abbate[1], G. Campanini[2], F. Rovida[2], L. Dossena[2], M. R. Capobianchi[1] and F. Baldanti[2]

1) *Laboratory of Virology, National Institute for Infectious Diseases INMI IRCCS 'L. Spallanzani', Rome and* 2) *S.S. Virologia Molecolare, S. C. Virologia e Microbiologia, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy*

Abstract

This study was aimed at establishing the genetic heterogeneity of influenza virus haemagglutinin (HA) gene quasi-species and the polymorphisms at codon 222, by application of ultra-deep pyrosequencing (UDPS) to respiratory samples from patients hospitalized for influenza A(H1N1)pdm09 infection, presenting with severe or moderate–mild disease. HA diversity was significantly higher in samples collected from patients with severe manifestations than in those from patients with moderate–mild manifestations (p 0.02). D222 polymorphism was detected in 40.7% of patients by UDPS, and in only 7.1% by Sanger sequencing. D222E, D222G, D222N and D222A were observed in 37.0%, 11.1%, 7.4% and 3.7% of patients, respectively; 10.7% of samples harboured more than two variants. The relative frequency of each single variant showed a wide range of intrapatient variation. D222G/N/A were detected, as either minor or predominant variants, only in severe cases, whereas D222E was equally represented in severe and moderate–mild infections. Other amino acid variants were observed at different positions within the analysed HA fragment. Consistent with higher heterogeneity, non-D222 variants were more frequently detected in severe cases than in moderate–mild cases. In addition, seven non-D222 mutations carried by minority variants, not previously described, were observed.

Keywords: D222 variants, haemagglutinin, influenza, polymorphisms, ultra-deep pyrosequencing
Original Submission: 5 March 2012; **Revised Submission**: 26 June 2012; **Accepted**: 29 June 2012
Editor: L. Kaiser
Article published online: 6 July 2012
Clin Microbiol Infect 2013; **19:** 668–673
10.1111/j.1469-0691.2012.03984.x

Corresponding author: M. R. Capobianchi, National Institute for Infectious Diseases 'L. Spallanzani', Via Portuense 292, 00149 Rome, Italy
E-mail: maria.capobianchi@inmi.it
M. Selleri and A. Piralla contributed equally to the study.

Background

The 2009 pandemic sustained by the influenza A(H1N1)pdm09 virus has been relatively mild, with moderate and self-limiting upper respiratory tract illness and no sign of pulmonary involvement in most cases. Nevertheless, in sporadic cases, influenza A(H1N1)pdm09 infections caused severe pneumonia and acute respiratory distress syndrome (ARDS), with patients requiring admission to intensive-care units and maximal life support measures [1].

Increased virulence of strains carrying the D222G/N substitutions in the receptor-binding glycoprotein of the virus (haemagglutinin (HA)) has been suggested, as these substitutions have been described in critically ill patients [2–9], whereas this issue is more controversial for the D222E substitution [2,10].

Influenza virus may be present in the infected host as a mixture of variants, referred to as quasi-species [11]. Next-generation sequencing (NGS) has emerged as an important tool with which to investigate minority variants in viral quasi-species for a number of viruses, including HIV, hepatitis C virus, and hepatitis B virus [12–15]. For influenza virus, the concomitant presence of different HA variants in clinical samples from patients infected with influenza A(H1N1)pdm09

has been detected with the metagenomic approach, based on different NGS platforms [16,17]. In addition, ultra-deep pyrosequencing (UDPS), based on the amplicon approach, has been used to evaluate the intrapatient heterogeneity of the HA gene, to detect the emergence of oseltamivir resistance [18].

In this study, UDPS was applied to respiratory samples from patients hospitalized for influenza A(H1N1)pdm09 infection, with the aim of establishing possible correlations between viral heterogeneity and the presence of variants (mainly focusing on position 222) with clinical presentations of the disease.

Materials and Methods

Patients

Thirty-two patients, previously screened by standard Sanger sequencing (Ss) for the presence of amino acid polymorphisms at codon 222 of the HA gene of influenza A(H1N1)pdm09, were considered, giving a total of 36 samples (28 nasal swabs (NSs) and eight bronchoalveolar lavages (BALs)). All specimens were collected before antiviral treatment initiation.

Detailed descriptions of patient selection criteria and of ethical aspects are included in Data S1. Briefly, patients were stratified according to the criteria proposed by Zarychanski *et al.* [19], as follows: (i) ten (31.3%) patients with severe influenza-like illness (ILI); (ii) 12 (37.5%) patients with moderate ILI; and (iii) ten (31.3%) patients with mild ILI.

According to clinical conditions, the patients were assigned to two groups: severe ILI (group A) and mild–moderate ILI (group B).

Sequencing

Nucleic acids from NS or BAL purification, reverse transcription reaction, viral load determination and Ss were performed as previously described [8]. For each sample, the amount of cDNA subjected to UDPS analysis corresponded almost exactly to the copies present in 1 mL of the starting material (Table 1).

UDPS was carried out with the 454 Life Sciences platform (GS-FLX; Roche Applied Science, Monza, Italy) as previously described [12], using titanium chemistry.

UDPS error rate estimation

To measure the accuracy of the UDPS, a plasmid clone containing the region of interest was sequenced in parallel by UDPS and by Ss. The plasmid clone was obtained from a patient sample by inserting the PCR amplicon into a pCR2.1-TOPO vector (Invitrogen™; Life Technologies, Monza, Italy). Differences between the two methods were considered to be GS-FLX sequencing errors.

GS Amplicon Variant Analyzer software (v.2.5.3; Roche Applied Science) was used to correct GS-FLX sequences and to identify the substitutions. Namely, the sequences obtained by UDPS were aligned with the Ss clone sequence, and for each position the error rate was estimated, distinguishing deletions, substitutions, and insertions. The calculated error rates were 0.032%, 0.035% and 0.069% for the different substitution types.

On the basis of previous experience [13], we considered as true variability all changes whose frequency was at least 0.21%, i.e. three times the insertion error rate, which was the highest frequency of procedural/experimental errors. In our experimental conditions, the number of cDNA templates undergoing UDPS and the resulting coverage (Table 1) were sufficiently high to rule out a possible influence on this sensitivity threshold, according to previous estimates [20].

The amplification conditions for UDPS and the bioinformatics analysis are described in Data S1.

Statistical evaluation

For most variables, descriptive statistics, such as median with interquartile range (IQR), and proportion (%), were calculated. The Mann–Whitney *U*-test and χ^2-tests (or Fisher's exact test when applicable) were used for univariate analysis, as appropriate. A two-tailed p-value of <0.05 was considered to be significant.

For calculation of the frequency of the HA substitutions, each substitution was individually counted.

Results

Overall, UDPS was applied to 27 of 32 (84.4%) patients, for whom sufficient amounts of viral RNA had been obtained to perform NGS in addition to Ss. Of these patients, eight had severe ILI (group A) and 19 had moderate–mild ILI (group B). For one patient from group B, paired samples (BAL and NS) were analysed. The median ages were 38.0 years (IQR 28.8–47.0 years) for group A and 43.0 years (IQR 25.0–57.0 years) for group B (non-significant difference). The median duration of illness at the time of sampling, calculated as number of days from the start of symptoms to sample collection, was significantly higher for group A than for group B (median of 5 days (IQR 4.5–6.5 days) vs. 2 days (IQR 1–3 days); p 0.03). At presentation, all patients of group A had ARDS, while patients of group B had fever plus at least one of the following symptoms: pneumonia, cough, dyspnea, rhinitis, myalgia and pharyngodynia. The

TABLE 1. Quasi-species haemagglutinin diversity and position 222 polymorphisms in 28 clinical samples from patients with severe (group A) and moderate–mild (group B) clinical presentations of influenza A(H1N1)pdm09 infection

Group	Patient ID	Sample type	Viral load[a] (log_{10} copies/mL)	Position 222 variant detected			Coverage[c]	Diversity[d]
				Sanger	UDPS[b]	Frequency of variant estimated by UDPS (%)		
A	1	BAL	6.01	E	E/D/G	98.93/0.80/0.27	7349	3.27
	2	BAL	5.75	D/N	D/N/G/E/A	34.03/43.82/19.06/2.28/0.65	4466	41.76
	3	BAL	7.11	D	D	100	16 100	0.99
	4	NS	5.44	D	D	100	11 340	1.38
	5	NS	4.37	D/G	D/G/N	65.79/33.18/1.00	14 231	19.29
	6	NS	5.47	E	E/D	99.63/0.37	7310	0.63
	7	NS	6.84	D	D	100	16 205	1.93
	8	NS	6.44	D	D	100	8615	0.93
	Median (IQR)		5.88 (5.45–6.74)					1.66 (0.95–15.29)
B	9	NS	4.86	D	D	100	21 257	1.06
		BAL	7.66	D	D	100	9651	0.57
	10	NS	8.24	D	D	100	15 446	0.69
	11	NS	6.67	E	E/D	99.57/0.43	14 176	0.86
	12	NS	5.84	D	D	100	13 118	1.22
	13	NS	6.74	D	D	100	15 881	0.29
	14	NS	7.07	D	D	100	12 416	0.52
	15	NS	6.86	E	E/D	99.60/0.40	5545	0.50
	16	NS	8.16	D	D	100	16 425	0.92
	17	NS	6.21	D	D	100	13 762	1.47
	18	NS	6.00	D	D	100	9020	4.06
	19	NS	5.62	E	E/D	99.58/0.42	16 045	2.62
	20	NS	8.17	E	E/D	99.37/0.63	15 806	0.41
	21	NS	8.32	D	D	100	19 536	0.44
	22	NS	8.61	D	D	100	13 802	0.84
	23	NS	7.58	E	E/D	99.39/0.61	15 791	0.41
	24	NS	7.36	E	E/D	99.34/0.66	18 026	0.29
	25	NS	7.26	D	D	100	11 170	0.42
	26	NS	8.92	E	E/D	99.34/0.66	22 042	18.10
	27	NS	8.92	D	D	100	5931	0.69
	Median (IQR)		7.31 (6.33–8.22)					0.69 (0.43–1.18)
p (A vs. B)[e]			0.01					0.02

BAL, bronchoalveolar lavage; IQR, interquartile range; NS, nasal swab; UDPS, ultra-deep pyrosequencing.
[a]Viral load also represents the number of cDNA copies subjected to UDPS, as there was 1 mL of starting material, and all of the extracted RNA was reverse transcribed and sequenced.
[b]The order of the variants is in accordance with their relative frequency in each patient.
[c]Number of complete reads obtained from each sample by UDPS.
[d]Diversity was calculated by the use of DNA distance (Phylip package), and is expressed as mean substitutions/site $\times 10^{-4}$.
[e]Calculated by Mann–Whitney *U*-test.

demographic and clinical characteristics of the 27 patients are shown in Table S1.

In Table 1, viral load values, HA quasi-species diversity, and polymorphisms, detected at position 222 in the 28 samples successfully analysed by UDPS, are reported, along with the results obtained with Ss. Median viral load values in group A were significantly lower than in group B (5.88 log_{10} copies/mL, IQR 5.45–6.74 vs. 7.31 log_{10} copies/mL, IQR 6.33–8.22; p 0.01). A total of 370 462 sequences from all samples were obtained by UDPS. The intrasample diversity in group A was significantly higher than that in group B (median (IQR): 1.66 (0.95–15.29) $\times 10^{-4}$ vs. 0.69 (0.43–1.18) $\times 10^{-4}$ substitutions per site, p 0.02); diversity was not significantly correlated with the duration of illness ($r^2 = 0.17$, p 0.073). The median diversity in NS samples was not significantly different from that in BAL samples (median (IQR): 0.85 (0.29–1.45) $\times 10^{-4}$ vs. 2.13 (0.68–32.14) $\times 10^{-4}$ substitutions per site; p 0.20).

Concerning position 222, a single variant, i.e. the reference D222, was detected by both UDPS and Ss in four patients from group A and in 13 patients from group B (including the patient with double sample type, i.e. patient no. 9). The concomitant presence of multiple variants was detected by UDPS in 11 of 27 patients (40.7%), four from group A and seven from group B. Eight patients harboured two variants (E/D), two patients harboured three variants (E/D/G and D/N/G), and one patient harboured five variants (D/N/G/E/A). There was high intrapatient variability in the frequency of each position 222 variant, which ranged between 0.27% (minority population) and 100% (pure population). In addition, the D222 reference was always detected in samples with variants, accounting for 0.37–65.79% of the overall quasi-species. D222E was the most frequent variant in both severe and moderate–mild patients, whereas D222G, D222N and D222A were present (eventually as minority quasi-species) only in samples from patients with severe

disease (3/8 in group A vs. 0/19 in group B, p 0.04). As expected, the efficiency of detection of multiple variants by Ss was much lower, as polymorphisms were detected in only two patients of group A (patients 2 and 5), where the intra-patient prevalence of variants, estimated by UDPS, was >19% (Table 1).

Besides the substitutions at codon 222, UDPS revealed several substitutions at additional positions in the analysed region, with respect to the reference strain (Table 2). S203T was always observed as >98% of the quasi-species components, consistent with the predominant clade 7 origin of Italian isolates [21]. Excluding S203T, a total of 27 non-D222 substitutions were detected, involving more frequently group A than group B patients (8/8 vs. 6/19, respectively, p 0.003). This was consistent with the increased heterogeneity in severe as compared with moderate–mild cases. All of these substitutions, with the exception of V199A and R205K, were present as minority variants (range of intrasample frequency: 0.22–1.97%). A stop codon was detected at position 188 in patient 11, as a minority variant (0.78%). Seven of the non-D222 substitutions had not been previously described (S203R, P218L, V220M, K239Q, G264S, T270A, and T277I; Table 2).

TABLE 2. Frequency of non-D222 substitutions detected by ultra-deep pyrosequencing in the patients with severe (group A) and moderate–mild (group B) clinical presentations; the substitutions that have never been reported in the literature are in bold

	Group A (*n* = 8)			Group B (*n* = 20)		
	n[a]	Frequency (%)	Patient no.	*n*[a]	Frequency (%)	Patient no.
Q188stop	–	–	–	1	0.78	11
D196G	1	0.84	4	–	–	–
V199A	–	–	–	1	95.47	18
S203I[b]	1	0.27	1	–	–	–
S203R	–	–	–	2	0.33; 0.38	19; 17
R205G	1	0.33	1	–	–	–
R205K	2	94.36; 99.64	2; 6	1	99.53	16
P218L	1	0.31	2	–	–	–
V220M	1	0.22	5	–	–	–
R221K	1	0.58	8	–	–	–
E235K	1	0.29	1	–	–	–
E235G	2	0.39; 0.27	1; 5	–	–	–
K239Q	1	0.28	5	–	–	–
T241K				1	0.68	9 (NS)
V249M	1	0.27	3	–	–	–
M257I	–	–	–	1	0.52	18
E258K	1	0.40	2	–	–	–
G264S	1	0.36	2	–	–	–
T270A	1	0.29	5	–	–	–
P271L	–	–	–	1	0.35	19
V272A	1	0.29	5	–	–	–
V272I	1	1.97	7	–	–	–
N276D	1	0.38	5	–	–	–
T277I	1	0.31	1	–	–	–

NS, nasal swab.
[a]*n*, number of patients showing the indicated substitution.
[b]The frequency of S203T, present as a majority variant in all of the samples, has been omitted.

Discussion

In this study, attention was paid to influenza A(H1N1)pdm09 HA quasi-species analysed by UDPS directly in clinical samples, with the aim of evaluating genetic heterogeneity and quantifying D222 variants present in viral quasi-species within each infected patient. Higher heterogeneity was observed in the samples collected from severe patients than in those collected from moderate–mild patients, and this difference was not dependent on sample type or the duration of infection.

Concerning the position 222 variants, the concomitant presence of multiple variants, including E/D, E/D/G, D/N/G and D/N/G/E/A, was revealed by UDPS in >40% of patients. These variants represented either predominant or minority components of viral quasi-species, and were not revealed by Ss when present at low frequency (i.e. <19%).

The presence of multiple D222 variants has been found in clinical samples by studies based on either standard population sequencing or a cloning/sequencing approach: Drews *et al.* [22] observed the D/N/G polymorphism in a severe patient, and Wang *et al.* [23] detected a mixture of D/N/G/S in both ICU and mild patients. Mixed viruses with D222G/A in patients with severe manifestations have also been reported in one Spanish patient [10].

NGS, based on the shotgun approach, was used in a study by Kuroda *et al.* [16], who reported D/N/G in a lung autopsy sample from a patient who died of viral pneumonia with influenza A(H1N1)pdm09 infection. In this study, variants carrying D222G/N were detected as minority components of viral quasi-species. More recently, Yasugi *et al.* [24] reported the presence of minority variants at position 222 in the HA from clinical samples obtained early during the influenza A(H1N1)pdm09 pandemic.

The present study has addressed for the first time, with an amplicon-based UDPS approach, the issue of genetic heterogeneity and polymorphisms at position 222 with respect to the severity of clinical presentation. The reference D222 was detected by UDPS in all patients, as a minor, relevant or even exclusive component of viral quasi-species. D222E appeared to be the most frequent variant, in both severe and moderate–mild patients, confirming the lack of association with disease severity previously reported [2].

In our series, the D222G/N variants (eventually present as minor components of viral quasi-species) were observed only in patients with a severe clinical presentation, in keeping with previous data [2,3,7,8].

Although no paired samples (NS and BAL) from severe patients were available in this study, cross-sectional data from previous studies indicate that these variants are more

frequent in the lower than in the upper respiratory specimens [3,9]. Several studies have indicated that D222G confers enhanced binding to α2–3-linked sialic acids, suggesting a greater ability to bind to the lung cells in the lower respiratory tract and cause an exacerbation of disease [7,25]. Furthermore, Belser *et al.* [26] have mutated an influenza A(H1N1)pdm09 virus to generate D222G, in order to study the role of D222G in pathogenesis and transmission in ferret and mouse models. Although the presence of this mutation did not cause a significant increase in pathogenesis and transmission in ferrets, it caused heightened virulence in mice, and exhibited enhanced replication in human respiratory cells.

The possible clinical significance of minor components of viral quasi-species is an attractive field of investigation, and may unravel new aspects of viral pathogenesis. Concerning influenza, enhanced cytopathogenicity for the lower respiratory tract, exerted by an even minute fraction of the replicating viral population, may lead to strong functional impairment of the respiratory epithelium, and to a more severe clinical presentation. In addition, the interplay of minority components of viral quasi-species with the cellular immune response may cause early viral adaptation and escape from immune control (as has been described to occur in primary HIV infection [27]), possibly leading to increased severity.

The presence of multiple variants in individual patients (and, possibly, in different body compartments), showing very diverse frequencies, seems not to be limited to position 222, but to be a general phenomenon, and supports the hypothesis of viral evolution within individual patients. The implications of these findings are particularly interesting from a pathogenetic and epidemiological standpoint, as they suggest that the emergence and the predominance of a more aggressive variant may be a single-patient-dependent phenomenon. Viral diversity was not correlated with the time interval between the start of symptoms and sample collection, so the increased diversity in group A could not be accounted for only by longer disease duration in this group (see Results). However, more information on this point may come in the future from the analysis of sequential samples from individual patients. In addition, future studies should be focused on the identification of factors that promote the evolution of influenza virus strains in single individuals. In fact, in other respiratory viral diseases (e.g. rhinovirus infections), the ability of the virus to invade the lower respiratory tract and lead to protracted infection is correlated with an impaired host immune response, rather than with viral genetic characteristics [28].

One limitation of the present study is the small number of analysed patients; despite this limitation, the results clearly show that UDPS allows the detection of variants present eventually as minute proportions of the overall quasi-species, evaluation of their contribution to the viral quasi-species in each single patient, and correlation of their presence with clinical severity.

The short region analysed represents an important drawback of studies performed with UDPS; in fact, different viral genome regions may differ substantially in the extent of variability, particularly for RNA viruses. In particular, HA represents one of the most variable influenza virus genome regions in both intrapatient and interpatient analyses [13,21], with a strong immune selection pressure, as indicated by the non-synonymous/synonymous substitution ratio, which is higher than in other genome segments [21]. Therefore, interpretation of the heterogeneity data obtained from UDPS analysis of a short HA fragment requires caution, and the conclusions cannot be simply extrapolated to the entire viral genome.

Despite this, we think that these result may represent a proof of concept and a starting point for further studies on the application of UDPS in the investigation of the role of (minor) variants of influenza virus quasi-species in disease severity.

Acknowledgements

We thank all the technical laboratory staff who contributed to the study.

Transparency Declaration

This work was partially supported by the Italian Ministry of Health (fondi Ricerca Corrente to National Institute for Infectious Diseases INMI IRCCS 'L. Spallanzani' and to Fondazione IRCCS Policlinico San Matteo), by the Istituto Superiore di Sanità (CCM 2009), by Regione Lombardia, by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 278433-PREDEMICS, and by Fondazione Cariplo (grant 2011). No authors have a commercial or other association that might pose a conflict of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Methods.

Table S1. Demographic and clinical characteristics of 27 patients with severe (group A) and moderate–mild (group B) clinical presentations of influenza A(H1N1)pdm09 infection.

References

1. Writing Committee of the WHO Consultation on Clinical Aspects of Pandemic (H1N1) 2009 Influenza. Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. *N Engl J Med* 2010; 362: 1708–1709.
2. Kilander A, Rykkvin R, Dudman SG, Hungnes O. Observed association between the HA1 mutation D222G in the 2009 pandemic influenza A(H1N1) virus and severe clinical outcome, Norway 2009–2010. *Euro Surveill* 2010; 15: pii19498.
3. Chen H, Wen X, To KK *et al.* Quasispecies of the D225G substitution in the hemagglutinin of pandemic influenza A(H1N1) 2009 virus from patients with severe disease in Hong Kong, China. *J Infect Dis* 2010; 201: 1517–1521.
4. Chan PK, Lee N, Joynt GM *et al.* Clinical and virological course of infection with haemagglutinin D222G mutant strain of 2009 pandemic influenza A (H1N1) virus. *J Clin Virol* 2011; 50: 320–324.
5. Miller RR, MacLean AR, Gunson RN, Carman WF. Occurrence of haemagglutinin mutation D222G in pandemic influenza A(H1N1) infected patients in the West of Scotland, United Kingdom, 2009–10. *Euro Surveill* 2010; 15: pii19546.
6. Mak GC, Au KW, Tai LS *et al.* Association of D222G substitution in haemagglutinin of 2009 pandemic influenza A (H1N1) with severe disease. *Euro Surveill* 2010; 15: pii19534.
7. Liu Y, Childs RA, Matrosovich T *et al.* Altered receptor specificity and cell tropism of D222G hemagglutinin mutants isolated from fatal cases of pandemic A(H1N1) 2009 influenza virus. *J Virol* 2010; 84: 12069–12074.
8. Baldanti F, Campanini G, Piralla A *et al.* Severe outcome of influenza A/H1N1/09v infection associated with 222G/N polymorphisms in the haemagglutinin: a multicentre study. *Clin Microbiol Infect* 2010; 17: 1166–1179.
9. Piralla A, Pariani E, Rovida F *et al.* Segregation of virulent influenza A(H1N1) variants in lower respiratory tract of critically ill patients during the 2010–2011 seasonal epidemic. *PLoS One* 2011; 6: e28332.
10. Ledesma J, Pozo F, Ruiz MP *et al.* Substitutions in position 222 of haemagglutinin of pandemic influenza A (H1N1) 2009 viruses in Spain. *J Clin Virol* 2011; 51: 75–78.
11. Ramakrishnan MA, Tu ZJ, Singh S *et al.* The feasibility of using high resolution genome sequencing of influenza A viruses to detect mixed infections and quasispecies. *PLoS One* 2009; 4: e7105.
12. Rozera G, Abbate I, Bruselles A *et al.* Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. *Retrovirology* 2009; 6: 15.
13. Abbate I, Vlassi C, Rozera G *et al.* Detection of quasispecies variants predicted to use CXCR4 by ultra-deep pyrosequencing during early HIV infection. *AIDS* 2011; 25: 611–617.
14. Ninomiya M, Ueno Y, Funayama R *et al.* Use of Illumina deep sequencing technology to differentiate hepatitis C virus variants. *J Clin Microbiol* 2012; 50: 857–866.
15. Margeridon-Thermet S, Shulman NS, Ahmed A *et al.* Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naïve patients. *J Infect Dis* 2009; 199: 1275–1285.
16. Kuroda M, Katano H, Nakajima N *et al.* Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS One* 2010; 5: e10256.
17. Bartolini B, Chillemi G, Abbate I *et al.* Assembly and characterization of pandemic influenza A H1N1 genome in nasopharyngeal swabs using high-throughput pyrosequencing. *New Microbiol* 2011; 34: 391–397.
18. Ghedin E, Laplante J, DePasse J *et al.* Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J Infect Dis* 2011; 203: 168–174.
19. Zarychanski R, Stuart TL, Kumar A *et al.* Correlates of severe disease in patients with 2009 pandemic influenza (H1N1) virus infection. *CMAJ* 2010; 182: 257–264.
20. Simen BB, Simons JF, Hullsiek KH *et al.* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis* 2009; 199: 693–701.
21. Valli MB, Meschi S, Selleri M *et al.* Evolutionary pattern of pandemic influenza (H1N1) 2009 virus in the late phases of the 2009 pandemic. *PLoS Curr* 2010; 2: RRN1149.
22. Drews SJ, Pabbaraju K, Wong S *et al.* Surveillance of autopsy cases for D222G substitutions in haemagglutinin of the pandemic (H1N1) 2009 virus in Alberta, Canada. *Clin Microbiol Infect* 2011; 17: 582–584.
23. Wang B, Dwyer DE, Soedjono M *et al.* Evidence of the circulation of pandemic influenza (H1N1) 2009 with D222D/G/N/S hemagglutinin polymorphisms during the first wave of the 2009 influenza pandemic. *J Clin Virol* 2011; 52: 304–306.
24. Yasugi M, Nakamura S, Daidoji T *et al.* Frequency of D222G and Q223R hemagglutinin mutants of pandemic (H1N1) 2009 influenza virus in Japan between 2009 and 2010. *PLoS One* 2012; 7: e30946.
25. Chutinimitkul S, Herfst S, Steel J *et al.* Virulence-associated substitution D222G in the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding. *J Virol* 2010; 84: 11802–11813.
26. Belser JA, Jayaraman A, Raman R *et al.* Effect of D222G mutation in the hemagglutinin protein on receptor binding, pathogenesis and transmissibility of the 2009 pandemic H1N1 influenza virus. *PLoS One* 2011; 6: e25091.
27. Henn MR, Boutwell CL, Charlebois P *et al.* Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 2012; 8: e1002529.
28. Tapparel C, Cordey S, Junier T *et al.* Rhinovirus genome variation during chronic upper and lower respiratory tract infections. *PLoS One* 2011; 6: e21163.